Infrastructure for Innovation via Data- and Model-Driven Design

Ken Koedinger
Professor of Human-Computer Interaction & Psychology
Carnegie Mellon University

Director of LearnLab
Co-leader of The Simon Initiative

PI of LearnSphere

Computer-Aided Personalized Education Workshop
Nov 12, 2015
• Do you know whether or not your students are learning (e.g., while listening to your lecture)?

• Do you know when you are learning (e.g., while listening to a lecture)?
Infrastructure should support an *engineering process*

- Should we "Build into tools the latest results of how people learn"?
  - No!

- Build tools for an *iterative engineering process*
  - Support using theory, *new* data, & modeling to design better learning
The Simon Initiative Vision

A data-driven virtuous cycle of learning science research and innovative educational practice causes demonstrably better learning outcomes for students from any background or place.
Why is a data infrastructure & engineering process so critical?

- We don’t know what we know
- We can’t see learning
- Principles of learning are not general
Adaptive Technology: Not new, but can do better

Not new:

e.g., Cognitive Tutors have provided adaptive instruction since the mid 80s
... iStart, Andes,

Not done: Plenty room for improvement

Quality of personalization depends on quality of model & data used to produce it
My current cell phone company charges me $14.95 per month for service and $.13 per minute. PPS Cellular Phone Company has offered me $15.00 worth of free calls a month if I switch, but the charge is $.39 per minute.

1. How many minutes of calls can I get from PPS Cellular Phone Company for $50? What is the cost from my current company for that number of minutes?
2. How many minutes of calls can I get from my current company for fifty dollars? What is the cost from PPS Cellular Phone Company for that number of minutes?
3. What is the cost from both companies for sixty minutes?

4. After how many minutes of calls will the cost for both companies be the same?
Loops of data-responsive adaptation

Step loop: Support student paths

Task loop: Select what’s needed

Design loop: Design for needs of all
Adapt to “hidden skills” that all students find difficult

Which is harder for algebra students?

*Story Problem*
As a waiter, Ted gets $6 per hour. One night he made $66 in tips and earned a total of $81.90. How many hours did Ted work?

*Word Problem*
Starting with some number, if I multiply it by 6 and then add 66, I get 81.90. What number did I start with?

*Equation*
\[ x \times 6 + 66 = 81.90 \]

Math educators say: story or word is hardest

Students: equations are hardest

[Graph showing percent correct for story, word, and equation problems.]

Expert blind spot!
Algebra teachers, especially, incorrectly think equations are easy

Collect data to check your assumptions!

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
<th>%Correct*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann is in a rowboat on a lake. She is 800 yards from the dock. She then rows for ( m ) minutes back towards the dock. Ann rows at a speed of 40 yards per minute. Write an expression for Ann's distance from the dock.</td>
<td>800 - 40m</td>
<td>40%</td>
</tr>
</tbody>
</table>

*Heffernan & Koedinger, 1997*
Elements of infrastructure for sharing innovations

What should we share?

- Data
- Analytic methods
- Explanatory models of learners
  - Models of cognition, metacognition, motivation, self-regulated learning, dialogue, social interaction
- Ed tech
  - generates data, uses models

Formal models of learning can make non-obvious predictions about hidden skills

- SimStudent combines
  - *Representation* learning (probabilistic grammars)
  - *Skill* learning mechanisms

- Novel prediction

<table>
<thead>
<tr>
<th>Step</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2x=12</td>
<td></td>
</tr>
<tr>
<td>6=3x</td>
<td></td>
</tr>
<tr>
<td>-x=5</td>
<td></td>
</tr>
<tr>
<td>-24=-4x</td>
<td></td>
</tr>
</tbody>
</table>
A community data infrastructure to support online learning improvement.

Existing Resources

Send us data, models, data-generating tools (or links to such)!
Some existing review articles

- **Data-driven ITS development**

- **Educational Data Mining**

- **Learning science results**
  - Pashler et al. (2007). Organizing Instruction and Study to Improve Student Learning (NCER 2007-2004).
  - Ambrose et al. (2010). *How learning works: Seven research-based principles for smart teaching*.

- **Toward a learning engineering process**
Thank you!

http://learnlab.org

http://cmu.edu/simon

http://learnsphere.org

Ken Koedinger
koedinger@cmu.edu